



International Journal of Research in Agronomy

E-ISSN: 2618-0618
P-ISSN: 2618-060X
© Agronomy
NAAS Rating (2026): 5.20
www.agronomyjournals.com
2026; 9(7): 919-927
Received: 19-05-2026
Accepted: 23-06-2026

Anilkumar Lalasing Chavan
Ph.D. (Genetics and Plant
Breeding), KSNUAHS,
Shivamogga, Karnataka, India

Anil HM
Ph.D. Scholar, Indian Agricultural
Research Institute, New Delhi,
India

Nimish Bhushan CN
Ph.D. Scholar, University of
Agricultural Sciences, Bangalore,
Karnataka, India

Corresponding Author:
Anilkumar Lalasing Chavan
Ph.D. (Genetics and Plant
Breeding), KSNUAHS,
Shivamogga, Karnataka, India

Genomic selection and multi-omics integration in rice breeding: Mechanistic insights, persistent bottlenecks, and the path to climate-resilient cultivars

Anilkumar Lalasing Chavan, Anil HM and Nimish Bhushan CN

DOI: <https://www.doi.org/10.33545/2618060X.2026.v9.i7l.6040>

Abstract

Rice breeding has moved, within roughly two breeding-programme generations, from phenotype-driven recurrent selection to genome-informed prediction, and is now moving again toward prediction models that draw on several molecular layers at once. This review develops a critical synthesis of that second transition: why genomic selection alone reaches a ceiling for complex, environmentally sensitive traits, and how transcriptomic, proteomic, metabolomic and phenomic data extend prediction accuracy by capturing regulatory and physiological information that marker genotype cannot supply. Genomic selection accuracy in rice depends more on training population design and genetic architecture than on the specific statistical model applied, a finding with direct consequences for how breeding resources should be allocated. Multi-omics evidence shows that hybrid performance, seed quality, pigment accumulation, salinity tolerance and yield heterosis are each shaped by regulatory and metabolic variation that sits outside what genomic markers alone resolve, and that the value of adding an omics layer scale with a trait's genetic complexity and environmental sensitivity rather than being uniform across the breeding target. The evidence base is markedly uneven across omics layers: genomic, transcriptomic, proteomic and metabolomic integration is comparatively mature, while epigenomic and microbiome-linked contributions to rice improvement remain thinly documented relative to their conceptual importance. Machine learning methods are structurally well suited to the high-dimensional, heterogeneous data multi-omics integration produces, but demonstrated deployment in operational rice breeding lags behind this structural opportunity, and claims of near-term autonomous or foundation-model-driven breeding pipelines exceed what the available evidence supports. Data integration standards, population representativeness, computational cost, interpretability and field-level validation are identified as the binding constraints on wider adoption, ahead of any remaining conceptual gap in genome-to-phenotype biology.

Keywords: Genomic selection, multi-omics integration, genomic prediction, phenomics, metabolomics, genotype-by-environment interaction, climate-resilient rice breeding

Introduction

Rice sustains more than half of the global population as a primary caloric source, and the crop's genetic improvement therefore carries consequences that extend well past the breeding nursery. Over the past two decades, pressure on breeding programmes has intensified: shifting rainfall patterns, expanding saline coastal zones, and increasingly erratic temperature extremes have compressed the timeline within which new cultivars must be delivered. Conventional pedigree breeding, built around recurrent phenotypic selection across multiple generations, was never designed to move at this pace. Marker-assisted selection improved precision for traits governed by one or a few large-effect loci, but its logic weakens once a trait's genetic architecture becomes diffuse, involving many small-effect variants that interact with environment^[1]. Genomic selection emerged directly from this limitation: rather than tracking a handful of markers linked to a target gene, it uses genome-wide marker information to estimate the breeding value of an individual before that individual is ever phenotyped in the field^[2, 3].

A genome sequence alone, however, is an incomplete predictor of phenotype. It specifies what a plant could become under some set of conditions, not what a specific plant actually becomes under the conditions a breeding trial imposes. Closing that gap is the shared ambition behind two

developments that this review treats as one connected story rather than two separate technologies: the maturation of genomic selection as a prediction framework, and the accumulation of transcriptomic, proteomic, metabolomic and phenomic evidence documenting how genomic potential is realized, or fails to be realized, in a given tissue, stage and environment ^[4, 5]. The remainder of this review builds that connection deliberately, tracing how each additional omics layer addresses a specific limitation of genome-only prediction, where the evidence for that value is strong, where it is thin, and what stands between current demonstration studies and routine deployment in applied rice breeding.

Evolution of Rice Breeding

The move from conventional breeding to genomics-informed breeding did not happen as a single event; it accumulated through successive technical increments, each removing one constraint on selection efficiency. Classical pedigree and pure-line breeding depended on multi-generation phenotypic evaluation, a slow process precisely because phenotype is the endpoint of an entire developmental and physiological cascade rather than a direct readout of genotype. Molecular markers shortened that cascade for traits with tractable genetic architecture: quantitative trait locus mapping and marker-assisted backcrossing let breeders track resistance genes or major quality loci with confidence, and this approach still underlies much of applied rice breeding for traits such as blast or bacterial blight resistance ^[1]. Its reach stops where genetic architecture becomes complex.

High-density SNP genotyping platforms changed the resolution at which breeders could observe the rice genome, moving selection from a few dozen markers to tens of thousands of genome-wide polymorphisms assayed at once ^[6], and genome-wide variation maps built from large germplasm panels made it possible to trace the actual selection signatures left behind by decades of breeding ^[7]. Whole-genome resequencing efforts within national breeding programmes, for example systematic surveys of Vietnamese rice germplasm, show that improvement-associated genomic regions are often shared across independently developed elite lines, implying that historical selection has already narrowed usable diversity at several loci relevant to yield and adaptation ^[8]. Comparable functional-genomics-driven breeding pipelines developed in India illustrate how sequence-based gene discovery and marker deployment were combined within one applied programme rather than treated as separate disciplines ^[9]. Against this backdrop, near-finished, gap-free indica genome assemblies removed a further bottleneck by giving breeders a reference precise enough to resolve structural variation that shorter-read assemblies had missed ^[10], while advances in sequencing chemistries lowered the cost of applying these tools at population scale ^[11]. The net effect, traced across the broader arc of rice genomic research, has been a steady widening of the trait space genomics can address, from single-gene qualitative resistance toward genuinely polygenic, environmentally sensitive traits such as yield and quality ^[12]. Genomic selection marks the point where this trajectory crosses a threshold: markers stop being used mainly to find genes and start being used, in aggregate, to predict performance directly.

Genomic Selection in Rice

Genomic selection reframes breeding-value estimation as a whole-genome regression problem. Instead of identifying individual causal variants, a genomic relationship matrix built from thousands of markers predicts the genomic estimated breeding value, or GEBV, of untested individuals based on their

genetic similarity to a phenotyped and genotyped training population ^[13]. The accuracy of that prediction is not a fixed property of the method; it is jointly determined by the genetic architecture of the trait, the size and relatedness structure of the training population, marker density, and the statistical model chosen to convert marker effects into a predicted value. This dependency was demonstrated directly in tropical indica breeding lines, where prediction accuracy varied substantially across traits and proved more sensitive to training population composition and size than to the specific choice of statistical model, whether ridge-regression BLUP, Bayesian variants, or reproducing kernel Hilbert space regression ^[2]. That finding carries an uncomfortable implication for practical breeding: investing further in exotic statistical machinery yields diminishing returns once the training population itself is poorly matched, in relatedness and environment, to the lines under selection.

Later work closed part of that gap from the genetic-architecture side rather than the modelling side. Embedding results from de novo genome-wide association analysis directly into the prediction model, so markers with independent evidence of a phenotypic effect are weighted rather than treated as exchangeable, improved prediction for several agronomic traits over standard genomic best linear unbiased prediction ^[14]. This matters biologically because additive genomic-relationship models implicitly assume genetic effects are spread diffusely and roughly equally across the genome; when a trait is instead shaped by a smaller number of moderate-to-large-effect loci layered on a polygenic background, ignoring that structure discards real information. Hybrid rice breeding adds a further complication, because the trait of interest, hybrid performance, is a property of a cross rather than of either parent alone, and heterotic performance can only be partly predicted from the additive value of each parental line ^[15]. Selecting parental combinations for testcross performance therefore benefits from models incorporating dominance and specific combining ability rather than additive relationship matrices alone.

Two further empirical patterns constrain how far genomic information alone can be pushed. First, models trained on markers lose considerable accuracy when trait expression depends strongly on genotype-by-environment interaction, and multi-environment trial data combined into a single prediction framework tend to outperform single-environment models for traits with pronounced environmental sensitivity ^[16]. Selecting markers that show consistent effect across environments, rather than using all available markers indiscriminately, further improves predictive ability for such traits ^[17]. Second, an explicit environmental covariate structure, capturing measured environmental variables alongside marker genotype, has been proposed as a way of reinstating the missing environmental dimension in both genome-wide association analysis and genomic prediction, treating environment as a quantifiable input rather than unmodelled noise ^[18]. These observations converge from different directions on the same conclusion: the genome sets the ceiling on what is predictable, but realized accuracy depends on how faithfully training data and model structure represent the environments and genetic backgrounds a breeding programme actually intends to select within. Dense phenomic data was proposed early as one practical route to narrowing that gap, on the logic that low-cost phenotypic covariates could supplement sparse or environmentally confounded genomic information without requiring an entirely new prediction framework ^[19].

Multi-Omics in Rice Breeding

No single omics layer captures the causal chain linking DNA

sequence to field performance, and this is a direct consequence of how biological information is transformed at each regulatory step rather than a rhetorical flourish. The genome sets the boundary of what is possible, but pan-genome analyses of diverse cultivated and wild rice accessions show that a single reference assembly systematically underrepresents real genic and structural variation, including presence-absence variants absent from any one individual genome and, by definition, invisible to marker panels built against a single reference [20]. Extending pan-genome resources to dozens of genetically diverse accessions confirms that a meaningful fraction of gene content is dispensable, lineage-specific, or hidden from standard variant-calling pipelines altogether, and some of this hidden variation is enriched near loci already implicated in domestication and improvement traits [21]. Genomic selection models built on markers genotyped against an incomplete reference therefore start from an incomplete accounting of genetic variation, one structural reason why genome-only prediction plateaus for some traits regardless of training population size.

Transcriptomic and proteomic layers add the dimension the genome cannot supply on its own: which encoded genes are actually expressed, in which tissue, and translated into functional protein. Full-length transcript-based proteogenomic annotation of the rice genome has revealed substantial numbers of previously unannotated or misannotated gene models and protein isoforms, refining the very gene set downstream association and prediction analyses depend on [22]. Proteomic surveys mapping enzymes onto their corresponding metabolic pathways demonstrated early that transcript abundance and protein abundance are not interchangeable proxies for pathway flux, since post-transcriptional and post-translational regulation intervenes between the two [23], a distinction that matters directly for traits governed by enzyme activity rather than gene presence, such as many aspects of grain composition. Proteomic and metabolomic profiling of biotic stress responses similarly shows that resistance phenotypes frequently trace to shifts in defence-related protein and metabolite pools not obviously predictable from resistance-gene genotype alone [24].

Metabolomics occupies a particularly informative position in this hierarchy because metabolite pools sit closest to the phenotype

among the molecular layers, integrating upstream genomic, transcriptomic and proteomic variation together with the plant's physiological and environmental state at sampling. High-coverage metabolomic profiling linked to predictive modelling of rice grain quality traits showed that metabolite-based models could recover quality variation marker-based models missed, implying some genetic control over these traits acts through metabolic intermediates not well tagged by the marker panel used [25]. Genome-wide association analysis of the metabolome itself, treating individual metabolite concentrations as quantitative traits, has mapped numerous metabolite-associated loci in rice populations that often do not coincide with previously known agronomic QTL, indicating a layer of metabolic genetic control conventional phenotype-based GWAS would not have detected [26]. Building on this logic, metabolomic profiles measured in parental lines have been used to predict hybrid yield with accuracy that in some comparisons matched or exceeded marker-based prediction, particularly for heterotic yield components [27], and metabolome-based prediction of yield heterosis more broadly has been proposed as a complementary route to identifying superior hybrid combinations before extensive field testing [28]. These results argue against treating metabolomics as a downstream validation tool; used at the parental or early-generation stage, it functions as a genuinely independent predictor.

Phenomics contributes a different kind of value: not a new molecular layer but a way of measuring the integrated phenotype non-destructively, repeatedly, and at a scale that keeps pace with genomic data volume. High-throughput, image-based non-invasive phenotyping applied to salinity responses in rice identified tolerance loci that overlapped only partially with loci known from destructive endpoint phenotyping, because continuous, dynamic measurement over the stress period captured components of the tolerance response, such as the trajectory of growth suppression and recovery, that a single endpoint measurement cannot [29]. A genomic selection model trained against endpoint phenotypes will, by construction, be blind to the genetic control of dynamic stress trajectories, and phenomic platforms are among the few practical routes to supplying that missing training signal at scale [19].

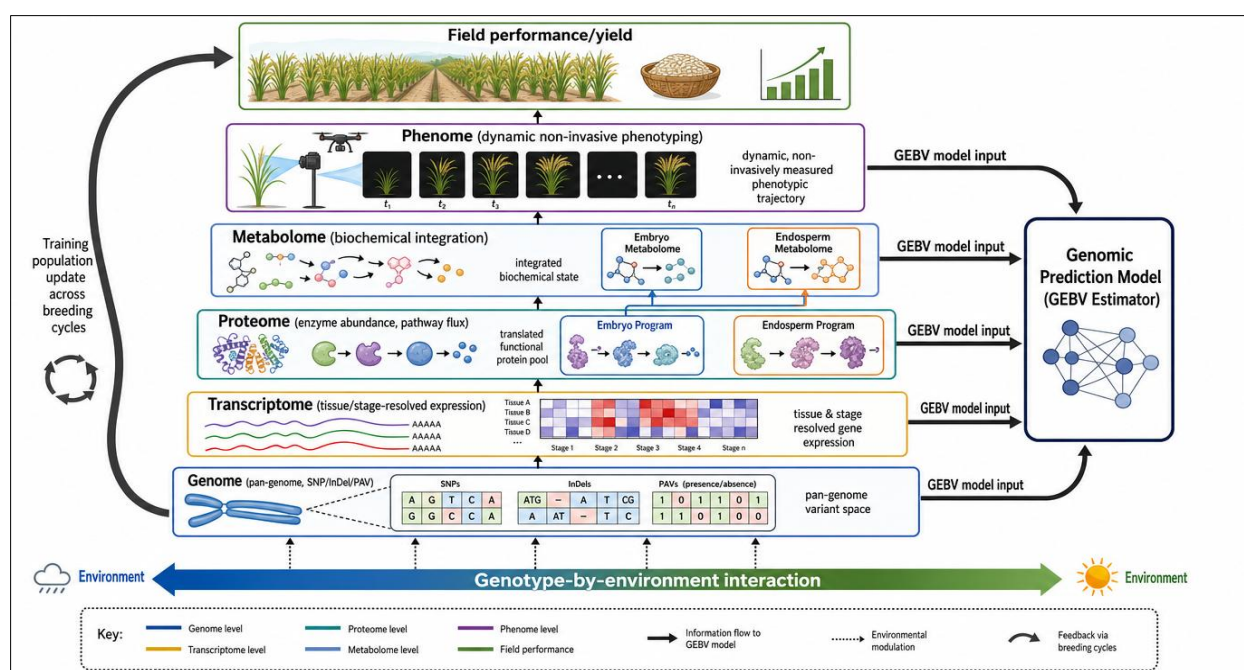


Fig 1: Tiered information flow from genome to phenotype in rice and points of multi-omics convergence with genomic prediction

Epigenomic and microbiome-linked omics belong conceptually to the same information hierarchy, mediating environmentally responsive gene regulation and plant-microbe interactions respectively, but they are comparatively underrepresented in the rice-specific literature synthesized here relative to the genomic, transcriptomic, proteomic and metabolomic layers discussed above; broader cereal-omics syntheses treat microbiome-associated trait variation as an emerging rather than established component of the multi-omics toolkit^[30]. This asymmetry is worth stating plainly rather than smoothing over: multi-omics rice breeding, as currently practiced, is considerably more mature for the molecular layers directly downstream of the genome than for the regulatory and ecological layers alongside it.

Integration of Genomic Selection and Multi-Omics

The scientific case for combining genomic selection with additional omics layers rests on a simple but often underappreciated asymmetry: genomic markers are stable across the plant's life cycle, while transcript, protein and metabolite abundances change continuously with developmental stage and environment, meaning they encode information about the realized, environment-specific expression of the genome that static marker genotype cannot. Incorporating multi-omic parental information, rather than parental marker genotype alone, into prediction models for hybrid rice measurably improved prediction of yield-related traits over marker-only models, with the improvement concentrated in traits where heterosis itself, rather than simple additive parental value, drives hybrid performance^[31]. Systematic comparison of alternative ways of combining omics layers for hybrid selection reinforces that no single integration strategy is universally optimal; the best-performing combination depends on which trait is being predicted, and metabolomic or transcriptomic layers sometimes outperform genomic markers for specific yield components even though genomic data remains the indispensable baseline layer^[32].

Tissue-specific multi-omics profiling illustrates why this trait-dependence exists mechanistically rather than as a statistical artefact. Comparative multi-omics characterization of rice embryo and endosperm tissue showed that these two seed compartments run substantially different transcriptional and metabolic programmes despite sharing the same genome, meaning seed-quality traits scored at the whole-grain level are composite outcomes of two only partially correlated tissue-specific processes^[33]. A genomic selection model predicting whole-grain quality without any tissue-resolved omics input implicitly averages over this heterogeneity, which caps achievable accuracy regardless of marker density. Comparable logic underlies multi-omics resources built around pigmented, nutritionally differentiated rice types, where pigment accumulation depends on coordinated transcriptional and metabolic changes across tissues that genome sequence alone does not resolve^[34], and multi-omics dissection of a rice population segregating for glycemic index and grain protein content has identified candidate genes and regions that only became interpretable once metabolomic and expression data were overlaid on the genomic association signal^[35]. Domestication-focused multi-omics analysis of black rice similarly traces pigment-pathway domestication to a combination of regulatory and structural genomic changes difficult to resolve from sequence variation alone^[36].

Abiotic stress tolerance offers a further illustration, and salt tolerance has become something of a proving ground for integrated approaches. Multi-omics dissection of salt tolerance mechanisms in rice consistently finds that transcriptional and

metabolic reprogramming under salinity involves coordinated shifts across ion transport, osmotic adjustment and antioxidant pathways that individual QTL rarely explain in full^[37], and dedicated reviews of integrative omics applied to rice salt response argue that transcriptomic and metabolomic data are necessary to resolve which of several candidate tolerance mechanisms actually operates in a given genetic background, since the same tolerance phenotype can arise through different physiological routes^[38]. This has a direct breeding consequence: two lines can show statistically indistinguishable salinity tolerance scores while relying on different molecular strategies, and a genomic selection model trained across both without omics-informed stratification risks blending two genetically distinct solutions into one, less interpretable, predictor.

At a more general systems level, formal frameworks for multi-omics integration in plant systems biology treat genomic selection as one node within a larger regulatory network model, in which prediction accuracy is expected to improve as additional omics layers reduce uncertainty in the genome-to-phenotype mapping^[39], and reviews cataloguing the broader status of multi-omics integration in crop improvement converge on a similar conclusion: the capacity to generate multiple omics layers has outpaced the statistical and computational infrastructure needed to combine them coherently^[40]. Targeted multi-omics interventions aimed at specific rice stress and trait problems report real gains from integration but consistently flag that these gains come from bespoke, trait-specific analytical pipelines rather than a generalizable integration framework transferring cleanly from one trait or population to the next^[41], and comparable multi-omics pipeline work in abiotic stress research more broadly reaches the same conclusion outside rice specifically, suggesting the bottleneck is generic to plant systems biology rather than particular to this one crop^[42].

Database infrastructure is not a peripheral concern here; it directly determines whether integration is even attempted. Purpose-built integrated omics knowledgebases for rice molecular breeding, bringing genomic, transcriptomic and phenotypic resources into a queryable common structure, exist precisely because the alternative, reassembling comparable data from separate repositories for every new analysis, is prohibitively slow^[43], and broader efforts to integrate omics databases across crop species report recurring obstacles: inconsistent metadata standards, incompatible identifier systems, and the absence of shared coordinate systems linking genomic position to expression and metabolite data^[44]. These are unglamorous problems, but they are the proximate reason multi-omics integration remains more common in single, well-resourced studies than in routine breeding pipelines.

Evidence that these integration principles are not idiosyncratic to rice comes from parallel work in other crops facing structurally similar problems. Multi-omics integration accelerated the molecular dissection of complex agronomic traits in common wheat through much the same logic of layering expression and metabolic data onto genomic association signals^[45], and comprehensive omics-based characterization of stress tolerance in sweet potato, a vegetatively propagated crop with reproductive biology very different from rice, converges on the same conclusion that single-omics analysis systematically underestimates the number of loci and pathways contributing to complex stress phenotypes^[46]. Multi-omics integration frameworks developed outside plant biology altogether, for instance layered genomic, epigenomic, transcriptomic, proteomic and metabolomic analysis of a complex human inflammatory skin condition, illustrate the same underlying statistical challenge of

combining heterogeneous, differently scaled data types into one coherent model, and analytical solutions proposed in that literature, particularly around handling missing modalities and weighting layers by informativeness, transfer directly to the plant breeding setting even though the biological system differs entirely^[47]. Emerging spatial multi-omics technologies, which preserve the physical location of a measurement within a tissue rather than homogenizing it, extend this integration logic further by adding a spatial coordinate to the existing multi-omics stack, a capability with obvious relevance to tissue-heterogeneous traits such as the embryo-endosperm distinction discussed above, although this technology has so far been demonstrated mainly outside cereal crops^[48].

Extending the same integrative logic to the association-mapping side of the pipeline, augmenting genome-wide association analysis with additional omics layers rather than relying on marker-phenotype association alone has been proposed as a way of recovering causal genes conventional GWAS misses because their marginal marker effect is too small to reach significance

alone but becomes interpretable once expression or metabolite data corroborate the candidate^[49]. Core-collection-based GWAS of agronomic traits in rice illustrates the scale of the underlying problem this is meant to solve: even carefully curated diversity panels return long lists of trait-associated regions with modest individual effect sizes, and distinguishing biologically meaningful associations from statistical noise in that list is precisely the task additional omics evidence is best positioned to assist^[50]. Translational genomics perspectives on crop breeding frame this whole progression, from single-marker genetics through genome-wide association to multi-omics integration, as a continuous broadening of the evidentiary base used to justify a selection decision rather than a series of separate technologies competing for adoption^[51], a framing consistent with how the multi-omics revolution in plant breeding more generally has been characterized: not as a replacement for genomic selection but as the mechanism by which its remaining blind spots are progressively closed^[5].

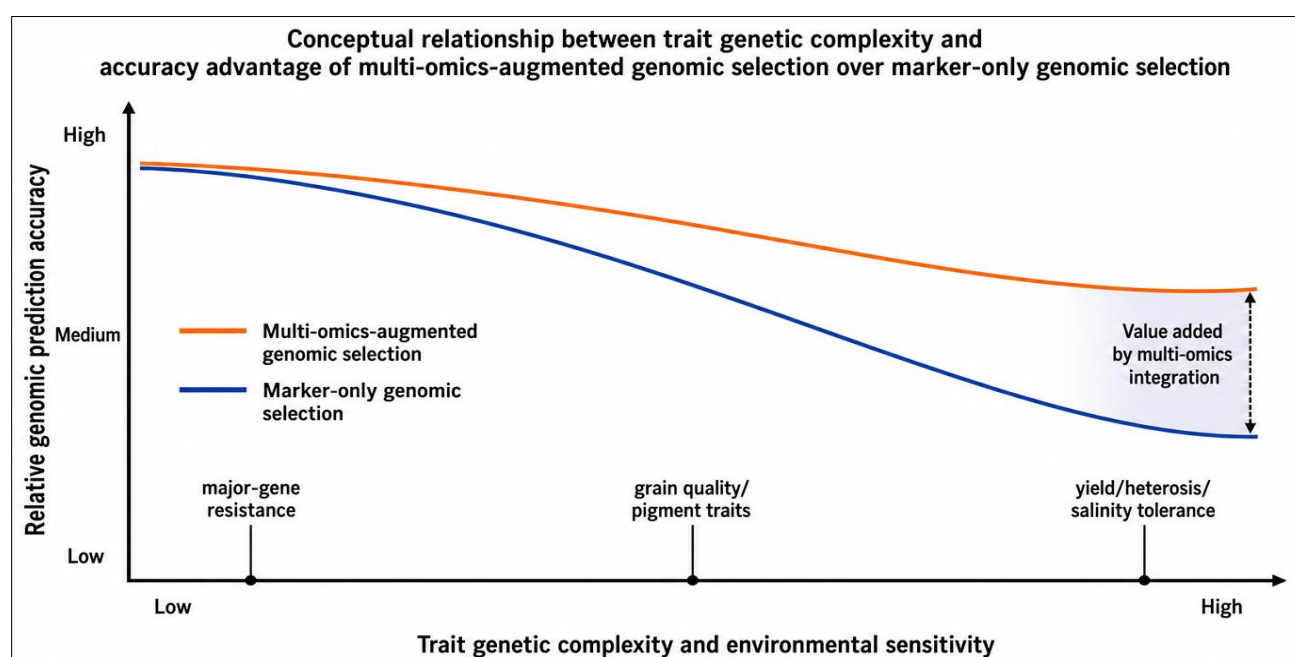


Fig 2: Conceptual relationship between trait genetic complexity and the accuracy advantage of multi-omics-augmented genomic selection over marker-only genomic selection.

Machine Learning and Artificial Intelligence in Genomic Prediction

The prediction models underlying genomic selection sit on a spectrum from classical mixed-model approaches to machine learning proper, and the evidence base examined here is informative mainly at the classical-to-intermediate end of that spectrum rather than at the deep-learning end. Direct comparison of ridge-regression BLUP, several Bayesian variants, and reproducing kernel Hilbert space regression across multiple agronomic traits in tropical rice found no single model dominating across traits, with gains from model choice consistently smaller than gains from improving training population design^[2]. This is a useful corrective to a common assumption elsewhere in the genomic prediction literature, that adopting a more flexible, heavily parameterized model is the most efficient route to higher accuracy; in the rice datasets actually examined, model flexibility mattered less than whether the training population adequately represented the genetic and environmental space of the target population. Reviews of

genomic prediction methodology for rice improvement treat model selection similarly, as a secondary lever useful for marginal gains once data quality and training population design are already adequate, rather than as the primary determinant of predictive success^[3].

Where the literature synthesized here does point toward a genuine role for more flexible, higher-capacity models is in multi-omics data fusion rather than in single-omics genomic prediction. Combining several omics layers to select optimal predictive combinations for hybrid rice required systematically comparing model architectures precisely because the relative informativeness of each omics layer for a given trait could not be assumed in advance, and this layer-weighting problem is structurally the setting in which more flexible, data-driven model classes are expected to outperform fixed linear combinations^[32]. Extending genome-wide association analysis with additional omics evidence to recover otherwise sub-threshold candidate genes similarly depends on a decision rule for how much weight corroborating expression or metabolite evidence should receive

relative to the marker association itself, a rule that becomes progressively harder to specify by hand as the number of integrated layers grows ^[49]. Metabolomics-based diagnostic breeding approaches, which use panels of metabolite biomarkers to classify or rank genotypes for downstream traits rather than measuring the trait directly, depend for their practical value on classification models capable of generalizing from a training panel of profiled lines to a much larger, unprofiled breeding population ^[52], and this is precisely the kind of pattern-recognition task where flexible statistical classifiers, machine learning included, offer a plausible advantage over simple threshold rules, even though the specific studies available here do not report head-to-head comparisons against deep-learning architectures.

Framed against these findings, claims that convolutional networks, graph neural networks or transformer-based foundation models are transforming rice genomic prediction should be treated as reasonable extrapolation from adjacent fields rather than as an established result within the literature examined here. The more defensible statement is narrower: multi-omics data fusion creates exactly the kind of high-dimensional, heterogeneous, partially redundant data structure that flexible machine learning methods are designed to handle, and the gap between that structural opportunity and demonstrated deployment in operational rice breeding remains real. Interpretability compounds this gap. A prediction from a linear mixed model can be decomposed into per-marker contributions a breeder can, in principle, relate back to known genomic regions; a prediction from a more flexible model trained across several omics layers is harder to decompose in the same way, and breeding programmes have historically been reluctant to act on predictions that cannot be explained to the people making final selection decisions in the field.

Applications in Climate-Resilient Rice

Translating these mechanistic and computational considerations into trait-specific breeding gains is clearest for salinity tolerance, where the largest body of integrated evidence in this reference set converges. Reviews tracking progress in salt-tolerant rice breeding note that, despite decades of QTL mapping effort, farmer-preferred high-yielding varieties with robust field-level salt tolerance remain comparatively scarce, and attribute this partly to the fact that most mapped tolerance QTL were identified under controlled, often single-stage stress screens that do not reproduce the fluctuating, multi-stage salinity exposure typical of coastal and delta rice fields ^[53]. Multi-omics characterization of the underlying tolerance mechanisms, combined with dynamic phenomic screening of the kind discussed earlier, offers a way of narrowing that gap by capturing tolerance as a trajectory across developmental stages rather than a single score, and by resolving which of several physiologically distinct tolerance strategies, ion exclusion, tissue tolerance or osmotic adjustment, predominates in a given genetic background ^[37, 38].

Disease resistance breeding presents a structurally different problem, since major resistance genes typically confer large, near-qualitative effects that classical marker-assisted selection already handles reasonably well; the harder remaining challenge is durability and the stacking of complementary resistance mechanisms as pathogen populations evolve. Integrating fine genetic mapping with complementary omics evidence for rice disease resistance has been proposed specifically to identify not just the resistance gene itself but the broader regulatory and metabolic context determining whether resistance holds up against divergent pathogen races in the field, since resistance genes deployed without attention to that context have repeatedly

been overcome within a few breeding cycles historically ^[54].

Grain quality and nutritional traits illustrate yet another pattern, one where metabolomics rather than genomics tends to carry the more direct predictive signal, because the trait being selected, starch structure, glycemic response, pigment content or protein composition, is itself a metabolic outcome rather than a structural or defence-related one. Multi-omics dissection of glycemic index and grain protein variation has identified candidate loci that only become interpretable alongside accompanying metabolite data, underscoring that grain-quality improvement is likely to benefit disproportionately from metabolomic integration relative to other trait classes ^[35], and the same logic extends to specialty pigmented rice types, where nutritionally and commercially valuable pigment accumulation is fundamentally a metabolic pathway trait ^[34]. Yield itself remains the hardest target precisely because it integrates contributions from nearly every layer discussed in this review, developmental timing, source-sink metabolic balance, stress resilience and pathogen exposure among them, consistent with the repeated observation that metabolome- and multi-omics-informed models add the most value for yield and yield-heterosis prediction relative to marker-only genomic selection, rather than for traits with simpler, single-pathway genetic control ^[27, 28].

Challenges

Several limitations already raised recur often enough across this literature to be treated as structural rather than incidental. Data integration remains the most frequently cited technical obstacle: omics platforms differ in throughput, dynamic range, missing-data patterns and the biological scale at which they operate, from single nucleotide to whole metabolite pool, and reconciling these differences within one statistical model is nontrivial even before biological interpretation begins ^[40]. High-dimensional, partially redundant multi-omics datasets increase the risk of overfitting a prediction model to the specific training population rather than to a generalizable biological signal, a risk that grows rather than shrinks as more omics layers are added unless training population size scales accordingly ^[3]. Missing data is not a minor nuisance here; it is close to unavoidable, because generating every omics layer on every individual in a large breeding population at every relevant developmental stage is rarely economically feasible, and most published multi-omics rice studies instead profile a genotyped subset intensively while leaving the remainder of the breeding population with genomic data alone ^[19].

Population and germplasm bias compounds this problem. Multi-omics resources have so far been built preferentially around well-characterized reference populations, elite breeding panels, or germplasm of particular commercial or nutritional interest, and prediction models trained on these resources do not automatically transfer to genetically distant breeding populations or to smallholder-relevant landraces that were never profiled at the omics level ^[34, 35]. Computational cost scales with this same imbalance: assembling, storing and querying multi-omics data at the volume required for operational breeding decisions demands infrastructure dedicated knowledgebase projects have had to build essentially from scratch for rice specifically ^[43], and cross-database integration efforts report that inconsistent identifiers and metadata standards across independently developed omics resources remain a persistent, largely unsolved obstacle to combining data across studies rather than within a single well-curated project ^[44].

Model interpretability and validation form a related but distinct constraint. A multi-omics-informed prediction can be statistically superior to a marker-only prediction in cross-validation and still fail to gain traction in an applied breeding programme if the basis

for that prediction cannot be communicated to the agronomists and breeders making final selection decisions, and this interpretability gap tends to widen as more omics layers and more flexible statistical methods are layered into a single pipeline ^[42]. Field deployment introduces a further validation problem: laboratory-stage multi-omics work does not fully anticipate: metabolite, transcript and even some phenomic measurements are sensitive to the precise growth conditions and developmental timing at sampling, so a predictive signal identified under one set of glasshouse or controlled-stress conditions may not reproduce cleanly across the variable multi-environment trial network operational breeding programmes actually rely on ^[41]. Emerging spatial multi-omics technology illustrates the same tension at a more advanced technical frontier: the capability to resolve omics signals within tissue architecture is scientifically compelling, but the throughput and cost currently associated with that resolution place it well outside what any but the best-resourced rice breeding programmes could deploy routinely ^[48]. None of these constraints is intractable, and none is close to solved either; this review's assessment is that they, rather than any remaining conceptual gap in genome-to-phenotype biology, are the binding constraint on faster adoption of multi-omics-assisted genomic selection.

Future Perspectives

Extrapolating cautiously from the patterns synthesized above, several directions look more like natural extensions of current evidence than speculative leaps. Pan-genome-informed genomic selection is the most immediately actionable of these: since pan-genome analyses have already demonstrated that reference-genome-based marker panels miss a nontrivial fraction of real genetic variation, including variation enriched near improvement-relevant loci ^[20, 21], incorporating pan-genome-derived variant calls directly into genomic relationship matrices is a logical next step for which the underlying variant catalogues already exist, even though prediction models built explicitly on pan-genome variation have not yet been widely benchmarked against standard single-reference approaches in the literature reviewed here. Continued reduction in sequencing cost, building on the near-complete assemblies and improved sequencing chemistries already achieved ^[10, 11], will make this more tractable at breeding-programme scale over coming breeding cycles rather than remaining confined to reference genome projects.

Spatial multi-omics, demonstrated so far mainly outside cereal crops ^[48], is a plausible but genuinely uncertain direction for rice: the tissue-resolved biology it could illuminate, echoing the embryo-endosperm heterogeneity already documented with conventional tissue-separated multi-omics ^[33], is real and relevant, but the cost and throughput required to apply spatial methods across a breeding-scale population rather than a handful of reference genotypes has not been demonstrated as feasible, and this review treats that gap as unresolved rather than assuming it will close on any particular timeline. Systems-biology-formalized integration frameworks, which model genomic selection as one node within a broader regulatory network rather than as a self-contained prediction task ^[39], point toward AI-assisted genomic prediction that weights each omics layer by its trait-specific informativeness rather than combining layers with fixed rules, extending the same layer-weighting logic already applied to multi-omics-augmented GWAS ^[49] into the prediction step itself. Whether this is best implemented through explainable machine learning architectures that preserve breeder-interpretable output, through Bayesian hierarchical extensions of existing genomic prediction models, or through some hybrid of the two is not resolved by the evidence surveyed here, and claims that fully

autonomous, foundation-model-driven breeding pipelines or digital-twin representations of individual breeding lines are close to operational readiness in rice go beyond what the supplied literature supports; these remain aspirational framings worth naming honestly rather than treating as an imminent, evidence-backed transition.

Conclusion

Genomic selection changed rice breeding by making prediction, rather than direct observation, the basis for selection decisions, and multi-omics data is best understood as the set of tools that make that prediction faithful to the real genotype-to-phenotype pathway rather than to an oversimplified additive-marker approximation of it. The evidence assembled here supports a fairly specific claim: single-omics genomic selection performs adequately for traits with tractable, moderately additive genetic architecture and works less well as trait complexity, environmental sensitivity, and physiological distance from the genome sequence increase, and multi-omics integration earns its computational and financial cost specifically in that harder regime, for yield, quality, and abiotic stress traits where genome sequence alone leaves too much unexplained variation on the table. What remains unresolved is less about biological understanding, which has advanced substantially across the studies reviewed, and more about the infrastructure, standardization, and validation pathways needed to move multi-omics-informed prediction out of well-resourced demonstration studies and into the routine operation of applied rice breeding programmes serving the varietal needs of a changing climate.

References

- Hasan N, Choudhary S, Naaz N, Sharma N, Laskar RA. Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes. *J Genet Eng Biotechnol*. 2021;19(1):128. <https://doi.org/10.1186/s43141-021-00231-1>
- Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redoña E, *et al*. Genomic selection and association mapping in rice (*Oryza sativa*): Effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet*. 2015;11(2):e1004982. <https://doi.org/10.1371/journal.pgen.1004982>
- Bartholomé J, Pra
- kash PT, Cobb JN. Genomic prediction: Progress and perspectives for rice improvement. In: *Methods in Molecular Biology*. Springer US; 2022. p. 569-617. https://doi.org/10.1007/978-1-0716-2205-6_21
- Yang Y, Saand MA, Huang L, Abdelaal WB, Zhang J, Wu Y, *et al*. Applications of multi-omics technologies for crop improvement. *Front Plant Sci*. 2021;12:563953. <https://doi.org/10.3389/fpls.2021.563953>
- Mahmood U, Li X, Fan Y, Chang W, Niu Y, Li J, *et al*. Multi-omics revolution to promote plant breeding efficiency. *Front Plant Sci*. 2022;13:1062952. <https://doi.org/10.3389/fpls.2022.1062952>
- Chen H, Xie W, He H, Yu H, Chen W, Li J, *et al*. A high-density SNP genotyping array for rice biology and molecular breeding. *Mol Plant*. 2014;7:541-553. <https://doi.org/10.1093/mp/ssT135>
- Xie W, Wang G, Yuan M, Yao W, Lyu K, Zhao H, *et al*. Breeding signatures of rice improvement revealed by a genomic variation map from a large germplasm collection. *Proc Natl Acad Sci U S A*. 2015;112(39):E5411-E5419.

- <https://doi.org/10.1073/pnas.1515919112>
9. Higgins J, Santos B, Khanh TD, Trung KH, Duong TD, Doai NTP, *et al.* Genomic regions and candidate genes selected during the breeding of rice in Vietnam. *Evol Appl.* 2022;15(7):1141-1161. <https://doi.org/10.1111/eva.13433>
 10. Agarwal P, Parida SK, Raghuvanshi S, Kapoor S, Khurana P, Khurana JP, *et al.* Rice improvement through genome-based functional analysis and molecular breeding in India. *Rice.* 2016;9(1):1. <https://doi.org/10.1186/s12284-015-0073-2>
 11. Du H, Yu Y, Ma Y, Gao Q, Cao Y, Chen Z, *et al.* Sequencing and de novo assembly of a near complete indica rice genome. *Nat Commun.* 2017;8:15324. <https://doi.org/10.1038/ncomms15324>
 12. Gu H, Liang S, Zhao J. Novel sequencing and genomic technologies revolutionized rice genomic study and breeding. *Agronomy.* 2022;12(1):218. <https://doi.org/10.3390/agronomy12010218>
 13. Yang X, Yu S, Yan S, Wang H, Fang W, Chen Y, *et al.* Progress in rice breeding based on genomic research. *Genes.* 2024;15(5):564. <https://doi.org/10.3390/genes15050564>
 14. Bhat JA, Ali S, Salgotra RK, Mir ZA, Dutta S, Jadon V, *et al.* Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Front Genet.* 2016;7:221. <https://doi.org/10.3389/fgene.2016.00221>
 15. Spindel JE, Begum H, Akdemir D, Collard BCY, Redoña ED, Jannink JL, *et al.* Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. *Heredity.* 2016;116:395-408. <https://doi.org/10.1038/hdy.2015.113>
 16. Cui Y, Li R, Li G, Zhang F, Zhu T, Zhang Q, *et al.* Hybrid breeding of rice via genomic selection. *Plant Biotechnol J.* 2020;18(1):57-67. <https://doi.org/10.1111/pbi.13170>
 17. Nguyen VH, Morante RIZ, Lopena V, Verdeprado H, Murori R, Ndayiragije A, *et al.* Multi-environment genomic selection in rice elite breeding lines. *Rice.* 2023;16(1):8. <https://doi.org/10.1186/s12284-023-00623-6>
 18. Bhandari A, Bartholomé J, Cao-Hamadoun TV, Kumari N, Frouin J, Kumar A, *et al.* Selection of trait-specific markers and multi-environment models improve genomic predictive ability in rice. *PLoS One.* 2019;14(5):e0208871. <https://doi.org/10.1371/journal.pone.0208871>
 19. Li X, Guo T, Wang J, Bekele WA, Sukumaran S, Vanous AE, *et al.* An integrated framework reinstating the environmental dimension for GWAS and genomic selection in crops. *Mol Plant.* 2021;14(6):874-887. <https://doi.org/10.1016/j.molp.2021.03.010>
 20. Sandhu KS, Merrick LF, Sankaran S, Zhang Z, Carter AH. Prospectus of genomic selection and phenomics in cereal, legume and oilseed breeding programs. *Front Genet.* 2022;12:829131. <https://doi.org/10.3389/fgene.2021.829131>
 21. Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, *et al.* Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat Genet.* 2018;50(2):278-284. <https://doi.org/10.1038/s41588-018-0041-z>
 22. Qin P, Lu H, Du H, Wang H, Chen W, Chen Z, *et al.* Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell.* 2021;184(13):3542-3558.e16. <https://doi.org/10.1016/j.cell.2021.04.046>
 23. Chen MX, Zhu FY, Gao B, Ma KL, Zhang Y, Fernie AR, *et al.* Full-length transcript-based proteogenomics of rice improves its genome and proteome annotation. *Plant Physiol.* 2020;182(3):1510-1526. <https://doi.org/10.1104/pp.19.00430>
 24. Koller A, Washburn MP, Lange BM, Andon NL, Deciu C, Haynes PA, *et al.* Proteomic survey of metabolic pathways in rice. *Proc Natl Acad Sci U S A.* 2002;99(18):11969-11974. <https://doi.org/10.1073/pnas.172183199>
 25. Vo KTX, Rahman MM, Rahman MM, Trinh KTT, Kim ST, Jeon JS. Proteomics and metabolomics studies on the biotic stress responses of rice: An update. *Rice.* 2021;14(1):29. <https://doi.org/10.1186/s12284-021-00461-4>
 26. Redestig H, Kusano M, Ebana K, Kobayashi M, Oikawa A, Okazaki Y, *et al.* Exploring molecular backgrounds of quality traits in rice by predictive models based on high-coverage metabolomics. *BMC Syst Biol.* 2011;5(1):176. <https://doi.org/10.1186/1752-0509-5-176>
 27. Gong L, Chen W, Gao Y, Liu X, Zhang H, Xu C, *et al.* Genetic analysis of the metabolome exemplified using a rice population. *Proc Natl Acad Sci U S A.* 2013;110(50):20320-20325. <https://doi.org/10.1073/pnas.1319681110>
 28. Xu S, Xu Y, Gong L, Zhang Q. Metabolomic prediction of yield in hybrid rice. *Plant J.* 2016;88(2):219-227. <https://doi.org/10.1111/tpj.13242>
 29. Dan Z, Chen Y, Zhao W, Wang Q, Huang W. Metabolome-based prediction of yield heterosis contributes to the breeding of elite rice. *Life Sci Alliance.* 2020;3(1):e201900551. <https://doi.org/10.26508/lsa.201900551>
 30. Al-Tamimi N, Brien C, Oakey H, Berger B, Saade S, Ho YS, *et al.* Salinity tolerance loci revealed in rice using high-throughput non-invasive phenotyping. *Nat Commun.* 2016;7(1):13342. <https://doi.org/10.1038/ncomms13342>
 31. Kaur B, Sandhu KS, Kamal R, Kaur K, Singh J, Röder MS, *et al.* Omics for the improvement of abiotic, biotic, and agronomic traits in major cereal crops: Applications, challenges, and prospects. *Plants.* 2021;10(10):1989. <https://doi.org/10.3390/plants10101989>
 32. Xu Y, Zhao Y, Wang X, Ma Y, Li P, Yang Z, *et al.* Incorporation of parental phenotypic data into multi-omic models improves prediction of yield-related traits in hybrid rice. *Plant Biotechnol J.* 2021;19(2):261-272. <https://doi.org/10.1111/pbi.13458>
 33. Wang S, Wei J, Li R, Qu H, Chater JM, Ma R, *et al.* Identification of optimal prediction models using multi-omic data for selecting hybrid rice. *Heredity.* 2019;123(3):395-406. <https://doi.org/10.1038/s41437-019-0210-6>
 34. Galland M, He D, Lounifi I, Arc E, Clément G, Balzergue S, *et al.* An integrated multi-omics comparison of embryo and endosperm tissue-specific features and their impact on rice seed quality. *Front Plant Sci.* 2017;8:1984. <https://doi.org/10.3389/fpls.2017.01984>
 35. Sedeek K, Zuccolo A, Fornasiero A, Weber AM, Sanikommu K, Sampathkumar S, *et al.* Multi-omics resources for targeted agronomic improvement of pigmented rice. *Nat Food.* 2023;4(5):366-371. <https://doi.org/10.1038/s43016-023-00742-9>
 36. Badoni S, Pasion-Uy EA, Kor S, Kim SR, Tiozon RN, Misra G, *et al.* Multiomics of a rice population identifies genes and genomic regions that bestow low glycemic index and high protein content. *Proc Natl Acad Sci U S A.* 2024;121(36):e2410598121. <https://doi.org/10.1073/pnas.2410598121>
 37. Zhou Z, Yang Z, Zhang Q, Wang X, Chen Q, Yu X, *et al.* Integrative multi-omics analysis reveals the domestication mechanism of black rice. *Natl Sci Rev.* 2025;13(10):nwaf497. <https://doi.org/10.1093/nsr/nwaf497>
 38. Dai L, Li P, Li Q, Leng Y, Zeng D, Qian Q. Integrated multi-omics perspective to strengthen the understanding of salt

- tolerance in rice. *Int J Mol Sci.* 2022;23(9):5236. <https://doi.org/10.3390/ijms23095236>
39. Ullah MA, Abdullah-Zawawi M, Zainal-Abidin R, Sukiran NL, Uddin MI, Zainal Z. A review of integrative omic approaches for understanding rice salt response mechanisms. *Plants.* 2022;11:1430. <https://doi.org/10.3390/plants11111430>
 40. Jamil IN, Remali J, Azizan KA, Nor Muhammad NA, Arita M, Goh HH, *et al.* Systematic multi-omics integration (MOI) approach in plant systems biology. *Front Plant Sci.* 2020;11:944. <https://doi.org/10.3389/fpls.2020.00944>
 41. Zhang R, Zhang C, Yu C, Dong J, Hu J. Integration of multi-omics technologies for crop improvement: Status and prospects. *Front Bioinform.* 2022;2:1027457. <https://doi.org/10.3389/fbinf.2022.1027457>
 42. Iqbal Z, Iqbal MS, Khan MIR, Ansari MI. Toward integrated multi-omics intervention: Rice trait improvement and stress management. *Front Plant Sci.* 2021;12:741419. <https://doi.org/10.3389/fpls.2021.741419>
 43. Roychowdhury R, Das SP, Gupta A, Parihar P, Chandrasekhar K, Sarker U, *et al.* Multi-omics pipeline and omics-integration approach to decipher plant's abiotic stress tolerance responses. *Genes.* 2023;14:1281. <https://doi.org/10.3390/genes14061281>
 44. Peng H, Wang K, Chen Z, Cao Y, Gao Q, Li Y, *et al.* MBKbase for rice: An integrated omics knowledgebase for molecular breeding in rice. *Nucleic Acids Res.* 2020;48(D1):D1085-D1092. <https://doi.org/10.1093/nar/gkz921>
 45. Chao H, Zhang S, Hu Y, Ni Q, Xin S, Zhao L, *et al.* Integrating omics databases for enhanced crop breeding. *J Integr Bioinform.* 2023;20(4):20230012. <https://doi.org/10.1515/jib-2023-0012>
 46. Zhang N, Tang L, Li S, Liu L, Gao M, Wang S, *et al.* Integration of multi-omics data accelerates molecular analysis of common wheat traits. *Nat Commun.* 2025;16(1):Article 57550. <https://doi.org/10.1038/s41467-025-57550-x>
 47. Ahmed S, Khan MSS, Xue S, Islam F, Ikram AU, Abdullah M, *et al.* A comprehensive overview of omics-based approaches to enhance biotic and abiotic stress tolerance in sweet potato. *Hortic Res.* 2024;11(3):uhae014. <https://doi.org/10.1093/hr/uhae014>
 48. Bratu D, Boda D, Caruntu C. Genomic, epigenomic, transcriptomic, proteomic and metabolomic approaches in atopic dermatitis. *Curr Issues Mol Biol.* 2023;45(6):5215-5231. <https://doi.org/10.3390/cimb45060331>
 49. Liu X, Peng T, Xu M, Lin S, Hu B, Chu T, *et al.* Spatial multi-omics: Deciphering technological landscape of integration of multi-omics and its applications. *J Hematol Oncol.* 2024;17(1):Article 96. <https://doi.org/10.1186/s13045-024-01596-9>
 50. Mangi N, Zhao Q, Li M, Li C, Yan L, Kong L, *et al.* Multi-omics-augmented GWAS for crop improvement: From mechanisms to breeding. *Mod Agric.* 2025;3(2):70032. <https://doi.org/10.1002/moda.70032>
 51. Zhang P, Zhong K, Zhong Z, Tong H. Genome-wide association study of important agronomic traits within a core collection of rice (*Oryza sativa* L.). *BMC Plant Biol.* 2019;19(1):295. <https://doi.org/10.1186/s12870-019-1842-7>
 52. Choi HK. Translational genomics and multi-omics integrated approaches as a useful strategy for crop breeding. *Genes Genomics.* 2019;41(2):133-146. <https://doi.org/10.1007/s13258-018-0751-8>
 53. Razzaq A, Wishart DS, Wani SH, Hameed MK, Mubin M, Saleem F. Advances in metabolomics-driven diagnostic breeding and crop improvement. *Metabolites.* 2022;12:511. <https://doi.org/10.3390/metabo12060511>
 54. Qin H, Li Y, Huang R. Advances and challenges in the breeding of salt-tolerant rice. *Int J Mol Sci.* 2020;21(21):8385. <https://doi.org/10.3390/ijms21218385>
 55. Ontoy JC, Ham JH. Mapping and omics integration: Towards precise rice disease resistance breeding. *Plants.* 2024;13(9):1205. <https://doi.org/10.3390/plants13091205>